

Combinatorics of The Interrupted Period.

Adrien Thierry *

April 25th, 2014

Abstract

This article is about discrete periodicities and their combinatorial structure. It describes the unique structure caused by the alteration of a pattern in a repetition. That alteration of a pattern could be “heard” as the disturbance that one can hear when a record is scratched and jumps.

Let x be a primitive word and x_1 be a proper prefix of x . Write $x = x_1x_2$ for a proper suffix x_2 of x . Let $W = x^{e_1}x_1x^{e_2}$ with $e_1 \geq 1, e_2 \geq 1, e_1 + e_2 \geq 3$.

Definition 1. Let \tilde{p} be the prefix of length $|lcp(x_1x_2, x_2x_1)| + 1$ of x_1x_2 and \tilde{s} the suffix of length $|lcs(x_1x_2, x_2x_1)| + 1$ of x_2x_1 . The factor $\tilde{s}\tilde{p}$ is the *core of the interrupt* of W .

Theorem 2. *Any factor of length $|x|$ of W containing W 's core of the interrupt is unique.*

Proof. Even though it is natural to define \tilde{p} as the prefix of length $|lcp(x_1x_2, x_2x_1)| + 1$ of x_1x_2 and \tilde{s} as the suffix of length $|lcs(x_1x_2, x_2x_1)| + 1$ of x_2x_1 when it comes to define the core of the interrupt, for the “clarity” of the proof, we will use $p = lcp(x_1x_2, x_2x_1)$ and $s = lcs(x_1x_2, x_2x_1)$ during all the proof. Deza, Franek, T. have shown that $lcs(x_1x_2, x_2x_1) + lcp(x_1x_2, x_2x_1) \leq |x| - 2$ when x is primitive (see [1]).

*McMaster University, Advol, adrien.thierry@gmail.com

Set $x = x_1x_2 = ps_p = p_s s = prs$ and $\tilde{x} = x_2x_1 = p'_s s = pr's$ for non-empty words p_s, s_p, p'_s, s'_p, r and r' .

Write $w = x_2x_1x_1x_2x_1x_2x_1x_2 = p'_s sp_s sps_p = pr'sprsprs$. During all the proof, the positions will refer to w (except when stated otherwise).

Set $i = |x| - |lcp(x_1x_2, x_2x_1)|$. Note that, any word of length $|x|$ starting at position $k, i \leq k \leq 2|x|$, as a cyclic shift of x , has 2 “natural” occurrences in w : one starting at position $k + |x|$, the other one at position $k + 2|x|$. Suppose, in order to derive a contradiction, that a factor v of w , of length $|x|$, and starting at a position $j, 0 \leq j < i$, has other occurrences in w . For each of these occurrences, there are two possibilities:

- either that occurrence starts at $k < |lcp(x_1x_2, x_2x_1)|$. But every word starting before $|lcp(x_1x_2, x_2x_1)|$ is a cyclic shift of x_2x_1 . By *synchronisation principle* (Ilie, [2]), \tilde{x} , and therefore x , are not primitive: a contradiction.
- or that occurrence starts at $k \geq |lcp(x_1x_2, x_2x_1)|$, (we only have to consider the cases $k < |lcp(x_1x_2, x_2x_1)| + |x|$).
 - If sp does not appear in r nor r' : Denote by r_p the first letter of r , by r_s the last one. Denote by r'_p the first letter of r' and by r'_s the last one. Then v contains $r'_s spr_p$ which does not appear in $(prs)^2$.
 - If sp appears in r : for each of its a^{th} appearance (except for the last one) consider the starting position i_a in $prsp$ of its factor p (that is each time I find sp I “mark” where the p is). Set $i_{v,s} = i - j$ the starting position of s in v and $i_{v,p} = i_{v,s} + |s|$ the starting position of p in v . Set $m = \max_{i_a} lcs(v[0...i_{v,p} - 1], (prsp)^2[0...i_a]) + 1$ and $M = \max_{i_a} lcp(v[i_{v,p}...end], (prsp)^2[i_a...end]) + 1$. The existence of an occurrence of $w[m...M]$ in x^2 would contradict the maximality of the lcs and lcp .
 - The symmetric of the proof completes the proof (i.e. working with the word $w' = x_2x_1x_2x_1x_2x_1x_1x_2$ first solves the cases where sp appears in r'

and then duplicates the result for the first part of W (which was not necessary (by translation)... *But the author finds it funny and beautiful !)).

□

Notes:

- By rotation, the result holds for any non-empty factor x_2 of $x = x_1x_2x_3$ with x_1 and x_3 possibly empty prefix and suffix of x and $W = x^{e_1}x_1x_3x^{e_2}$. This is what makes me say “alteration of a pattern”, in a deletion way, in the abstract. For clarity, the theorem is stated in its form.
- In the case $lcs(x_1x_2, x_2x_1) + lcp(x_1x_2, x_2x_1) = |x| - 2$, any word of length $|x| - 1$ and containing sp will have $e_1 + e_2$ occurrences.
- Any factor of length $|x|$ of W not containing the core of its interrupt has $e_1 + e_2$ occurrences.
- This is a birthday theorem.

Thanks to my supervisors Antoine Deza and Franya Franek for helpfull discussions and advices and to Alice Heliou for proof reading a first version of this article.

References

- [1] A. Deza, F. Franek, and A. Thierry. How many squares can a string contain ? *submitted to publication*, 2014.
- [2] Lucian Ilie. A simple proof that a word of length n has at most \sqrt{n} distinct squares. *Journal of Combinatorial Theory, Series A*, 112(1):163 – 164, 2005.